Group-Testing with at most c Tests for Finite c and $c \to \infty$

by

Satindar Kumar and Milton Sobel[1]

University of Wisconsin at Milwaukee

University of Minnesota

Technical Report No. 146

University of Minnesota
Minneapolis, Minnesota

October 1970

[1]Research supported by NSF Grant GP-11021.

1. Introduction.

In group-testing one tests a batch of x units and observes one of the two mutually exclusive results

i) either all x are good or

ii) at least one is defective. (We don't know which ones and for $x > 1$ we don't know how many.)

It may happen that we can schedule a fixed number of tests for a certain time period (say, for the day) and we would like to maximize the number of units classified in this period. If each test takes the same amount of time a restraint on the total time is equivalent to a restraint on the total number of tests. We assume (at the outset) an infinite binomial population, each element of which is defective with probability p and good with probability $q = 1 - p$; all these are mutually independent. For convenience, we assume that the units are ordered and we consider only those procedures which have the "first-come, first-served" property, i.e., if the unit U_1 is in front of unit U_2 then U_1 is classified before or at the same time as U_2. We actually invoke a related stronger assumption that a defective set, i.e., a set of size $m \geq 2$ known to contain one or more defectives, must always be analyzed (without mixing it with binomial units) before returning to the binomial or H-situation in which $m = 0$. The procedure R_T, introduced in Section 2 below, satisfies these assumptions.

The corresponding group-testing problem for a fixed total number of units N, (and an unlimited number of tests available), which attempts to minimize the expected number of tests required to classify all N units, was treated in [4] and [5]. Previous formulations for group-testing were considered by Dorfman [1] and Sterrett [8]. A multinomial version of the group-testing problem, in which the categories best, second best, etc., are ordered and the outcomes are nested, was considered in [2]. Hypergeometric group-testing, where the total number of defectives is known in advance, was considered in [6]. References [3], [4], [5], [6] and [7] all use a common dynamic programming approach that is also used in this paper. An information-theoretic approach is treated in the appendix of [4], which is also appropriate for the present paper. A third procedure $R_{2,1}$, which is also appropriate to our formulation with $N = \infty$, is considered in [5]. Some comparisons of the results for these procedures with the procedure R_T of this paper are made in Section 5. A major result of this paper is the proof that the basic procedure R_T approaches the procedure $R_{2,1}$ as $c \to \infty$.

2. Recursion Formulas for Procedure R_T.

Let $U(m, c) = U_D(m, c|R_T)$ denote the expected number of units classified under procedure R_T if we have at most c tests remaining and we start with a defective set of size m, which is known to contain at least one defective. For $m = 0$ we have no defective set, i.e., all the unclassified units are binomial, and we write $U(c) = U_B(c|R_T)$ if there are at most c tests remaining. Using x as the integer number of units for the next test we can now write the recursions that define procedure R_T.

For $c \geq 1$ and $m \geq 2$

$$(2.1) \qquad U(c) = \max_{x=1,2,\dots} \{q^x[x + U(c-1)] + (1-q^x)U(x, c-1)\},$$

$$(2.2) \qquad U(m, c) = \max_{1 \le x < m} \{(\frac{q^x - q^m}{1 - q^m})[x + U(m-x, c-1)] + (\frac{1 - q^x}{1 - q^m})U(x, c-1)\}.$$

The boundary conditions for $m = 1$ and/or $c = 0$ are

(2.3) $\qquad U(0) = 0,$

(2.4) $\qquad U(m, 0) = 0$ for $m \geq 2$, and

(2.5) $\qquad U(1, c) = 1 + U(c)$ for $c \geq 0$.

The probabilities that enter into (2.1) are obvious; those that enter into (2.2) were derived in [4] and [5] and need not be repeated here.

For $x = 1$ and any c, the right side of (2.1) simplifies to $1 + U(c-1)$ and for $c = 1$ this is 1. Suppose, for example, $c = 1$. Then we have to compare 1 with $2q^2$, $3q^3$, $4q^4$ etc., and this leads to the following result (with the maximizing x shown on the right)

$$
(2.6) \qquad U(1) = \begin{cases} 1 & \text{for } 0 < q < 3^{-1/3} \; ; \; x = 1 \\ 3q^3 & \text{for } .693\ldots < q < 3/4 \; ; \; x = 3 \\ 4q^4 & \text{for } .750 < q < 4/5 \; ; \; x = 4 \\ 5q^5 & \text{for } .800 < q < 5/6 \; ; \; x = 5 \\ \ldots & \text{for } .833\ldots < q \; ; \; x \geq 6 \\ \ldots & \ldots \quad \ldots \end{cases}
$$

where $.693\ldots = 3^{-1/3}$; for $c = 1$ we never take $x = 2$ in the binomial situation.

To obtain the value of $U(2)$ we first need to obtain the following results (with the maximizing x shown on the right)

$$(2.7) \qquad U(2, 1) = \frac{1 + 2q}{1 + q} \text{ for all } q \; ; \; x = 1,$$

$$(2.8) \qquad U(3, 1) = 1 \qquad \text{for all } q \; ; \; x = 1,$$

$$(2.9) \qquad U(4, 1) = 1 \qquad \text{for all } q \; ; \; x = 1,$$

$$
(2.10) \qquad U(5, 1) = \begin{cases} 1 & \text{for } 0 < q < q_{1,2}^{(5,1)} \; ; \; x = 1, \\ \dfrac{2q^2(1-q^3)}{1 - q^5} & \text{for } q_{1,2}^{(5,1)} < q < 1 \; ; \; x = 2, \end{cases}
$$

where $q_{1,2}^{(5,1)} = .848\ldots$ is the root of $1 + q - q^2 - q^3 - q^4 = 0$. From these we find that $U(2)$ can be written as

$$U(2) = \begin{cases} 1 + U(1) = 2 & \text{for } 0 < q < q_{1,2}^{(2)} \; ; x = 1 \\ 1 + q + q^2 U(1) & \text{for } q_{1,2}^{(2)} < q < q_{2,4}^{(2)} \; ; x = 2 \\ 1 + 3q^4 + q^4 U(1) & \text{for } q_{2,4}^{(2)} < q < q_{4,5}^{(2)} \; ; x = 4 \\ \ldots & \ldots \qquad ; x \geq 5 \end{cases} \tag{2.11}$$

where $q_{1,2}^{(2)} = .618\ldots$ is the root of $1 - q - q^2 = 0$, where $q_{2,4}^{(2)} = .786\ldots$ is the root of $1 - 3q^3 + 4q^5 - 4q^7 = 0$, and where $q_{4,5}^{(2)} = .834\ldots$ is the root of $3 - 4q + 6q^6 - 6q^7 = 0$. For $c = 2$ the integer $x = 3$ is never used in the binomial situation.

We now prove by induction

Lemma 1:

For $q < q_{1,2} = \frac{1}{2}(\sqrt{5} - 1)$ and for any integer $c \geq 1$ we always take $x = 1$ in the binomial situation.

Proof:

The result holds by direct computation for $c = 1$. For $c \geq 2$ we have to compare $1 + U(c-1)$ and

$$\begin{aligned} & q^2[2 + U(c-1)] + (1-q^2)U(2, c-1) \\ &= q^2[2 + U(c-1)] + q(1-q)[1 + U(1, c-2)] + (1-q)U(1, c-2) \\ &= q + q^2 + q^2U(c-1) + (1-q^2)[1 + U(c-2)] \\ &= 1 + q + q^2U(c-1) + (1-q^2)U(c-2). \end{aligned} \tag{2.12}$$

Hence the dividing point between $x = 1$ and $x = 2$ is at the value of q where

$$(2.13) \qquad U(c-1) = \frac{q}{1-q^2} + U(c-2).$$

By the induction hypothesis, for $q < q_{1,2}$ we have $U(c-1) = 1 + U(c-2)$ and this shows that the root $q_{1,2}$ of $1 - q - q^2 = 0$ or $q_{1,2} = \frac{1}{2}(\sqrt{5} - 1)$ is the required dividing point betwen $x = 1$ and $x = 2$.

We can also show by induction with the help of (2.13) that in a small interval of the form $(q_{1,2}, q_{1,2} + \varepsilon)$ the derivative of $U(c)$ increases with c for $c \geq 2$. It follows that $x = 2$ is preferable to $x = 1$ in this interval and $q_{1,2} = .618\ldots$ is the dividing point between one-at-a-time sampling and group-sampling (with $x \geq 2$) in the binomial situation for $c \geq 2$.

Remark 1:

It should be noted that the above recursive scheme can also be used for any finite number of units N. Let $n \leq N$ denote the current number of unclassified units and write the recursion with $U(c|n)$ and $U(m, c|n)$; the only change is in (2.1) where the range of x is from 1 to n. Since $m \leq n$ the solution of (2.2) never depends on n and the solution of (2.1) depends on n only when q is sufficiently large so that the required x-value (under procedure R_T for $N = \infty$) is larger than n.

Remark 2:

In writing (2.2) we have assumed that whenever a (non-empty) defective set is present we always take the next test group solely from the defective set. In [5] it is shown that for $m = 2$ and q sufficiently close to 1 an improvement to the procedure R_1 can be found by a so-called mixing procedure, provided the 2 units in the defective set are identifiable. Moreover this mixing procedure is "first-come, first-served." The same counterexample holds for procedure R_T with $m = 2$ and $c \geq 1$. Take 2 other unclassified (binomial) units and consider these $n = 4$ units. Label the 2 units in the defective set as d_1 and d_2. For the mixing routine we use d_1 and the 2 binomial units for the next group test. For $q \to 1$ the expected number of units classified per test approaches $n/2 = 2$ for the mixing procedure, while the corresponding result for procedure R_T approaches 3/2. Hence the mixing procedure is better than R_T for $n \geq 4$ and q close to one. It follows that for specified values of q, all the mixing subroutines introduced in [5] and [7] can be used to improve R_T for q close to one. However it is shown in [7] that procedure R_1 is at least 90 percent efficient for all q when compared to the improved procedure and it is conjectured that a similar result holds for procedure R_T (cf. Section 5).

Remark 3:

The function $U(c)$ increases approximately linearly with c and tends to infinity as $c \to \infty$. Hence for purposes of comparison the limit of $U(c)/c$ as $c \to \infty$ will be of interest. This limit, which is the expected number of units classified per test, is comparable with the reciprocal of the limiting value studied in [5]. That limiting value, $C(q; R)$, is the expected number of tests required per unit classified. For example, for two different procedures, $R = R_{2,1}$ and $R = R_{0,1}$ the common result $c(.99; R) = .08105\ldots$ was obtained in [5]. It is shown below that the limit of $U(c)/c$ as $c \to \infty$ for procedure R_T is equal to $1/.08105 = 12.338$ to 5 significant figures for $q = .99$.

3. The Symmetric Problem with a Fixed Total Number of Tests.

Another procedure R_{ST}, based on a fixed total number of tests, is for the problem in which each test on x units gives 3 possible results:

i) Either all x are good or

ii) all x are defective or

iii) at least 1 good and at least 1 defective are present.

Let $U_{DG}(m, c) = U_{DG}(m, c|R_{ST})$ denote the expected number of units classified in at most c tests under procedure R_{ST} if we start with a DG-set (i.e., a set containing at least 1 defective and at least 1 good unit) of size $m \geq 2$. Similarly, let $U_D(m, c)$ (resp., $U_G(m, c)$) denote the same if we start with a set of size $m \geq 2$ containing at least 1 defective unit (resp., at least 1 good unit). For $m = 0$ these all reduce to a binomial situation and we write $U_B(c)$ for the binomial case with at most c tests remaining.

Here we use the same symbol m to denote the size of a 'defective set', a'good set' and a'mixed set'; this does not lead to confusion since we never have more than one of these sets at the same time.

For $c \geq 1$ and $m \geq 2$ the recursion formulas for R_{ST} are given by

$$(3.1) \qquad U_B(c) = \max_{x=1,2,\ldots} \{(p^x + q^x)[x + U_B(c-1)] + (1 - p^x - q^x)U_{DG}(x, c-1)\},$$

$$(3.2) \qquad U_{DG}(m, c) = \max_{1\leq x\leq m-1} \{(\frac{q^x - q^m}{1-p^m-q^m})[x + U_D(m-x, c-1)] + (\frac{p^x - p^m}{1-p^m-q^m})[x+U_G(m-x,c-1)]$$

$$+ (\frac{1-p^x-q^x}{1-p^m-q^m})U_{DG}(x, c-1)\},$$

$$(3.3) \qquad U_D(m, c) = \max_{1\leq x\leq m} \{\frac{p^x}{1-q^m}[x + U_B(c-1)] + (\frac{q^x-q^m}{1-q^m})[x + U_D(m-x, c-1)]$$

$$+ (\frac{1-p^x-q^x}{1-q^m})U_{DG}(x, c-1)\},$$

$$(3.4)\qquad U_G(m, c) = \max_{1\le x\le m} \{\frac{q^x}{1-p^m}[x + U_B(c-1)] + (\frac{p^x-p^m}{1-p^m})[x + U_G(m-x, c-1)] + (\frac{1-p^x-q^x}{1-p^m})U_{DG}(x, c-1)\}.$$

The boundary conditions are

$$(3.5)\qquad U_B(0) = 0 = U_{DG}(m, 0) = U_D(m, 0) = U_G(m, 0) \quad \text{for} \quad m > 1,$$

$$(3.6)\qquad U_D(1, c) = U_G(1, c) = 1 + U_B(c) \quad \text{for} \quad c \geq 0,$$

$$(3.7)\qquad U_D(0, c) = U_G(0, c) = U_B(c). \quad \text{for} \quad c \geq 0.$$

Remark 1:

The function $U(c)$ is symmetric in q about $q = 1/2$ and hence Lemma 1 above does not hold for procedure R_{ST}.

Remark 2:

It is easy to show that we never take $x = 1$ in a binomial situation, i.e., when $m = 0$. We do this by showing that $U_B(c) \neq 1 + U_B(c-1)$ and the inequality $>$ must then hold since $x = 1$ is always an available strategy. Apply the scheme for $U_B(c-1)$ to the first $c - 1$ tests for the $U_B(c)$ situation. At the end of these $c - 1$ tests we find ourselves in a $U_B(1)$ situation or in a $U_D(m, 1)$ situation for some m, with positive probability for both. Since $x = 2$ (for $c = 1$) gives $2(p^2 + q^2) \geq 2 - 4pq \geq 1$ and $x = 1$ gives $p + q = 1$, we find that $U_B(1) > 1$ for $p \neq 1/2$. Since $x = 1$ is always an available strategy, we also have $U_D(m, 1) \geq 1$. Hence for $p \neq 1/2$ we end up with more than $1 + U_B(c-1)$ for the value of $U_B(c)$ and this proves our result.

If $c \geq 2$ then it can be shown that even for $p = 1/2$ we would not take $x = 1$ in the binomial situation. In fact, if $x = 1$ is preferable to $x = 2$, then

$$(3.8) \qquad 1 + U_B(c-1) > (p^2 + q^2)[2 + U_B(c-1)] + 2pq[2 + U_B(c-2)]$$

or

$$(3.9) \qquad U_B(c-1) - U_B(c-2) > \frac{1}{2pq} .$$

If $x = 1$ is preferable to $x = 2$ for $U_B(c)$ then essentially the same argument as above shows that $x = 1$ will also be preferable for $U_B(c-1)$. (If not, we can get an improvement.) Hence $U_B(c-1) = 1 + U_B(c-2)$. Putting this in (3.9) the resulting inequality is false for all q and this proves the result.

Remark 3:

Since the negation of (3.9) holds and $U_B(0) = 0$ we obtain an upper bound for $U_B(c)$ by summing, i.e.,

$$(3.10) \qquad U_B(c) \leq \frac{c}{2pq}$$

which shows that $U_B(c)$ remains under a linear function of the form bc.

4. Asymptotic Analysis of Procedure R_T.

In this section we consider the asymptotic structure of procedure R_T as $c \to \infty$. It is shown that the procedure R_T approaches the procedure $R_{2,1}$ defined and studied for the assembly line case (i.e., when the original number of units $N = \infty$) in section 6 of [5]. We start with the basic assumption of the existence of a constant b such that for large c

$$(4.1) \qquad U(c) \sim bc,$$

where $b = b(q)$ can depend on q. This assumption is justified by the fact that it leads to explicit expressions for $U(m, c)$ in (4.14) for q large and for b in (4.18) which satisfy the basic recursion formulas (2.1) through (2.5). A numerical comparison of the results obtained for R_T with those obtained by the limiting procedure R_∞ $(= R_{2,1})$ for $q = .90$, $.95$ and $.99$ will be made in the next section.

To illustrate the method of getting R_∞ we first use (4.1) in (2.5); we also investigate which x yields a maximum in (2.2) for $m = 2, 3, \ldots$. Some results obtained (with the maximizing x shown on the right) are

$$U(1, c) \sim 1 + bc \qquad \text{for all } q \tag{4.2}$$

$$U(2, c) \sim \frac{1 + 2q}{1 + q} + (c-1)b \qquad \text{for all } q;\ x = 1 \tag{4.3}$$

$$U(3, c) \sim \frac{1 + 2q + 3q^2 + b}{1 + q + q^2} + (c-2)b \qquad \text{for all } q;\ x = 1 \tag{4.4}$$

$$U(4, c) \sim \begin{cases} \dfrac{1 + 2q + 3q^2 + 4q^3 + (2+q)b}{1 + q + q^2 + q^3} + (c-3)b & \text{for } 0 < q < q_{2,3};\ x = 1 \\[2ex] \dfrac{1 + 2q + 3q^2 + 4q^3}{1 + q + q^2 + q^3} + (c-2)b & \text{for } q_{2,3} < q < 1;\ x = 2, \end{cases} \tag{4.5}$$

where $q_{2,3} = .755\ldots$ is the(unique) root of $1 - q^2 - q^3 = 0$. An exact expression for $U(m, c)$ for $q > q_{m-1,m}$ is obtained below for all m; note that $q \;\; < q \;\; = .819$ so that our result below need only agree with the second expression in (4.5).

Using (4.2) in the right side of (2.1) for $x = 1$ and (4.3) for $x = 2$, we obtain the trial values U_1, U_2 for U

$$(4.6) \qquad U_1 = 1 + b(c-1)$$

$$(4.7) \qquad U_2 = 1 + q + b(c-2+q^2).$$

Equating these, we obtain at the dividing point between $x = 1$ and $x = 2$

$$(4.8) \qquad b = \frac{q}{1 - q^2} .$$

From (2.1) with bc on the left and $x = 1$ on the right, we obtain

$$(4.9) \qquad b = 1.$$

Hence from (4.8) and (4.9) the dividing point between $x = 1$ and $x = 2$ is the (unique) root $q_{1,2}$ of

$$(4.10) \qquad 1 - q - q^2 = 0,$$

and the value of b for $0 < q < q_{1,2}$ is 1 by (4.9); this result is consistent with lemma 1 above. Again with bc on the left side of (2.1) and $x = 2$ on the right we obtain

$$(4.11) \qquad b = \frac{1 + q}{2 - q^2} \quad \text{for } q_{1,2} < q < q_{2,3}.$$

The value of $q_{2,3}$ is obtained by comparing U_2 and U_3 and is found to be the (unique) root of $1 - q^2 - q^3 = 0$. An explicit expression for b for any q-value is found below. Furthermore we show that for all x the dividing point between x and $x + 1$ in the binomial situation under procedure R_∞ is the (unique) root (denoted by $q_{x,x+1}$) of

$$(4.12) \qquad 1 - q^x - q^{x+1} = 0.$$

For any $m \geq 2$ we define $r = r(m)$ and $d = d(m)$ by writing

$$2^r \leq m < 2^{r+1} \quad \text{and} \quad d = 2^{r+1} - m > 0. \tag{4.13}$$

We wish to show by induction

<u>Theorem 1:</u>

For q close to 1 (more precisely, for $q > q_{m-1,m}$)

$$\begin{aligned} U(m, c) &\sim (c-r-1)b + \frac{1}{p} + \frac{b(1-q^d) - mq^m}{1 - q^m} \\ &= \frac{1 + 2q + \ldots + mq^{m-1}}{1 + q + \ldots + q^{m-1}} + b\{c - r - 1 + \frac{1 - q^d}{1 - q^m}\} \end{aligned} \tag{4.14}$$

and the value of x that maximizes the right side of (2.2) is given by

$$x = x_{max} = \begin{cases} 2^{r-1} & \text{for } 2^r \leq m < 3\cdot 2^{r-1} \quad (\text{or } d > 2^{r-1}) \\ m - 2^r & \text{for } 3\cdot 2^{r-1} \leq m < 2^{r+1} \ (\text{or } d \leq 2^{r-1}). \end{cases} \tag{4.15}$$

<u>Proof:</u>

We consider two cases according as $d > 2^{r-1}$ or $d \leq 2^{r-1}$.

In Case 1 we have $2^r \leq m < 3\cdot 2^{r-1}$ and $d > 2^{r-1}$. By (4.15) $x = 2^{r-1}$, $d_x = x$, $r_x = r - 1$. Then $2^{r-1} \leq m - x = 3\cdot 2^{r-1} - d < 2^r$, $d_{m-x} = d - x$ and $r_{m-x} = r - 1$. Hence by the induction hypothesis (4.14) for x and $m - x$, we obtain for the right side of (2.2) for $x = 2^{r-1}$

$$\begin{aligned} &\left(\frac{q^x - q^m}{1 - q^m}\right)[x + (c-r-1)b + \frac{1}{p} + \frac{b(1-q^{d-x}) - (m-x)q^{m-x}}{1 - q^{m-x}}] \\ &+ \left(\frac{1 - q^x}{1 - q^m}\right)[(c-r-1)b + \frac{1}{p} + \frac{b(1-q^x) - xq^x}{1 - q^x}] \\ &= (c-r-1)b + \frac{1}{p} + \frac{b(1-q^d) - mq^m}{1 - q^m} = U(m, c) \end{aligned} \tag{4.16}$$

as required by (4.14). The proof that any value of $x > 2^{r-1}$ always leads to a smaller result and that any $x < 2^{r-1}$ leads to a smaller result for $q > q_{m-1,m}$ is postponed.

In Case 2 we have $3 \cdot 2^{r-1} \leq m < 2^{r+1}$ and $d \leq 2^{r-1}$. By (4.14) $2^{r-1} \leq x = m - 2^r < 2^r$, $d_x = 2^r - x$ and $r_x = r - 1$. Then $m - x = 2^r$, $d_{m-x} = 2^r$ and $r_{m-x} = r$. Since $x = m - 2^r$, we also have that $m - 2x = 2^{r+1} - m = d$. From the induction hypothesis (4.14) for x and $m - x$, we obtain for the right side of (2.2) for $x = m - 2^r$

$$(4.17) \qquad \left(\frac{q^x - q^m}{1 - q^m}\right)\left[x + (c-r-2)b + \frac{1}{p} + \frac{b(1-q^{m-x}) - (m-x)q^{m-x}}{1 - q^{m-x}}\right]$$

$$+ \left(\frac{1 - q^x}{1 - q^m}\right)\left[(c-r-1)b + \frac{1}{p} + \frac{b(1-q^{m-2x}) - mq^m}{1 - q^m}\right]$$

$$= (c-r-1)b + \frac{1}{p} + \frac{b(1-q^d) - mq^m}{1 - q^m} = U(m, c),$$

as required by (4.14). The proof that any smaller or larger value x yields a smaller result than (4.17) is postponed.

The x-values used above (and the corresponding dividing points) are exactly the same as those used for the G-situation (i.e., when $m \geq 2$) in the procedure R_1 in [4] and in the procedure $R_{2,1}$ in [5]. The x-values for the H-situation (and their corresponding dividing points given by (4.12)) are shown below to be exactly the same as those used in procedure R_2 in [4] and in procedure $R_{2,1}$ in [5]. Thus the procedure $R_{2,1}$ of [5] has all its x-values and dividing points in common with R_∞ and is therefore identical with R_∞.

Applying (4.14) for $U(x, c)$ in (2.1) for any x, we obtain an expression for $b = b_x$. The same expression holds for all q, if the appropriate

$x = x(q)$ under procedure R_∞ is inserted. Let $r' = r_x$ and $d' = d_x$ denote the r and d-values for x. Then from (2.1) we obtain

$$b_x = \frac{1 - q^x}{p[(r'+1)(1-q^x) + q^{d'}]} . \tag{4.18}$$

We now use (4.18) to prove that the root of (4.12) is the correct dividing point between x and $x + 1$ under procedure R_∞. Let d'' and r'' denote the d and r-values for $x + 1$. Consider two cases according as $r'' = r'$ or $r'' = r' + 1$. Under Case 1, $d'' = d' - 1$ and equating b_x and b_{x+1}, we obtain

$$\begin{aligned} 0 &= (1-q^x)[(r'+1)(1-q^{x+1}) + q^{d'-1}] - (1-q^{x+1})[(r'+1)(1-q^x) + q^{d'}] \\ &= 1 - q^x - q(1-q^{x+1}) = p(1-q^x-q^{x+1}). \end{aligned} \tag{4.19}$$

Under Case 2, $r'' = r' + 1$, $d' = 1$ and $d'' = x + 1$. Equating b_x and b_{x+1}, we obtain

$$\begin{aligned} 0 &= (1-q^x)[(r'+2)(1-q^{x+1}) + q^{x+1}] - (1-q^{x+1})[(r'+1)(1-q^x) + q] \\ &= (1-q^x)(1-q^{x+1}) + q^{x+1}(1-q^x) - q(1-q^{x+1}) = p(1-q^x-q^{x+1}). \end{aligned} \tag{4.20}$$

This completes the proof that (4.12) gives the correct dividing points in the H-situation under procedure R_∞.

It is easy to show that the q-values $q_{j,j+1}$ ($j = 1, 2,\ldots$) form an increasing sequence which approaches one; we omit the proof. For any given $q = q_0$ we now define the correct integer $x = x(q)$ under procedure R_∞ in the H-situation by finding the interval $[q_{x-1,x}, q_{x,x+1}]$ that includes q_0. With $q = q_0$ and $x = x(q)$ inserted, (4.18) gives the correct $b = b(q)$ under procedure R_∞ for all values of q.

As an illustration, suppose $q = .99$. From Table VII of [4] we find that $x = 69$ between $q = .9899$ and $q = .9901$. Hence $r' = 6$, $d' = 2^7 - 69 = 59$ and (4.18) becomes

$$(4.21) \qquad b(q|x = 69) = \frac{1 - q^{69}}{p[7(1-q^{69}) + q^{59}]} \quad \text{for} \quad .9899 < q < .9901.$$

In particular for $q = .99$ we obtain $b = 12.338\ldots$. This value of b and its reciprocal .08105 were obtained by some other method in Table I of [5] for procedures $R_{2,1}$ and $R_{0,1}$ (cf. p. 138 of [5]). For $q = .95$ and .90 we use the same table to find that $x = 14$ and 7, respectively, and we obtain

$$(4.22) \qquad b = \frac{1 - q^{14}}{p[4(1-q^{14}) + q^2]} \quad \text{for} \quad .9499 < q < .9533,$$

$$(4.23) \qquad b = \frac{1 - q^7}{p[3(1-q^7) + q]} \quad \text{for} \quad .8987 < q < .9116.$$

The numerical values of b for $q = .95$ and .90 are given in Table 1 and also agree with the results in [5].

It is of some interest to point out that $b = b(q)$ in (4.18) is a continuous function of q that starts at 1 for $q < .618\ldots$ and $b \to \infty$ as $q \to 1$. To show this we first note that for q close to 1 the value of x is approximately given by

$$(4.24) \qquad 0 = 1 - q^x - q^{x+1} \sim 1 - 2q^x \quad \text{or} \quad x \sim [\log_2 y]^{-1}$$

where $y = 1/q$. Since $d \leq x$ and $q^x \sim 1/2$, we obtain for (4.18) for q close to one

$$(4.25) \qquad b \sim \frac{\frac{1}{2}}{p[(r+1)\frac{1}{2} + q^d]} \sim \frac{1}{p \log_2 x} \sim \frac{-y}{(y-1)\log_2 \log_2 y},$$

which is easily shown to approach ∞ as y (or q) $\to 1$ from above (below).

It is plausible to conjecture that $b(q)$ in (4.18) is a strictly increasing function of q but this has not been shown.

Appendix to Section 4

To complete our induction proof we have to show that the trial values are maximized when we take x as in (4.15). For Case 1 suppose $x = 2^{r-1} + f$ where $0 < f \le 2^{r-2}$ and consider two cases (1A and 1B) according as $m > 2^r + f - 1$ or $m \le 2^r + f - 1$. For Case 1A we have $2^{r-1} \le m - x < 2^r$ so that $r_{m-x} = r - 1$ and $d_{m-x} = 2^r - m + x = d + x - 2^r = d - x + 2f$. Also $r_x = r - 1$ and $d_x = 2^{r-1} - f = x - 2f$. To show that the new value is smaller we have to show that

$$(4.26) \qquad q^x(1 - q^{d-x+2f}) + 1 - q^{x-2f} \le 1 - q^d$$

or that

$$(4.27) \qquad q^d(1-q^{2f}) \le q^{x-2f}(1-q^{2f}).$$

This holds for all q since $0 < d_{m-x} = d - x + 2f$ and $f \ge 0$.

For Case 1B we have $2^{r-2} \le m - x < 2^{r-1}$ so that $r_{m-x} = r - 2$ and $d_{m-x} = 2^{r-1} - m + x = d - 2x + 3f$. Also $r_x = r - 1$ and $d_x = 2^{r-1} - f = x - 2f$. We now have to show that

$$(4.28) \qquad q^x - q^m + q^x - q^{d-x+3f} + 1 - q^{x-2f} \le 1 - q^d$$

or that

$$(4.29) \qquad q^x(1-q^{m-d}) \le q^{3f} + q^{2x-2f-d} - 2q^{2x-d}.$$

Suppose $m = 2^r + g$ where $0 \leq g \leq f - 1 < x$, so that $d = 2^{r+1} - m = 2^r - g$ $= 2x - 2f - g$ and (4.29) can be written as

$$(4.30) \qquad q^{x-g}(1-q^{2g}) + q^{2f}(1-q^{f-g}) \leq (1-q^{f+g}) + q^{f+g}(1-q^{f-g}).$$

Dividing by $1 - q$ and noting that $x > f > g$, the inequality now follows from a term-by-term comparison.

We now consider Case 1 with $f > 2^{r-2}$; the case in which $f \geq 2^{r-1}$ is omitted. For Case 1 with $2^{r-2} < f < 2^{r-1}$ we have $3 \cdot 2^{r-2} < x < 2^r$ so that $r_x = r - 1$ and $d_x = 2^{r-1} - f = x - 2f$. Under Case 1 we have $2^r \leq m < 3 \cdot 2^{r-1}$ and hence $0 < m - x < 3 \cdot 2^{r-2}$. We distinguish two cases (1C and 1D) according as $m - x \geq 2^{r-1}$ or $m - x < 2^{r-1}$.

Under Case 1C we have $r_{m-x} = r - 1$ and $d_{m-x} = 3 \cdot 2^{r-1} - m + f = x - g$. We have to show that

$$(4.31) \qquad q^x(1-q^{x-g}) + 1 - q^{x-2f} \leq 1 - q^{m-2g}$$

or, equivalently, that

$$(4.32) \qquad q^{m-2g}(1-q^{2f}) \leq q^{x-2f}(1-q^{2f}).$$

Replacing g by $m - 2^r$ we find that (4.32) holds if and only if $m \leq 3 \cdot 2^{r-1} + f$, which is true for Case 1C.

For Case 1D we use $m - x < 2^{r-1}$ to show that $2^r + g - (2^{r-1} + f) < 2^{r-1}$ or $g < f$. Assuming $2^{r-j-1} \leq m - x < 2^{r-j}$ for some j $(1 \leq j < r - 1)$, we have $r_{m-x} = r - j - 1$ and $d_{m-x} = 2^{r-j} - m + x = d'$ (say). We have to show that

$$(4.33) \qquad j(q^x - q^m) + q^x(1 - q^{d'}) + 1 - q^{x-2-f} \leq 1 - q^{m-2g}.$$

We first show that the number of terms on the right (after dividing by $1 - q$) in (4.33) is greater than that on the left, i.e., that

$$(4.34) \qquad j(m-x) + 2^{r-j} - (m-x) + x - 2f < m - 2g$$

or, equivalently, after replacing f by $x - 2^{r-1}$ and g by $m - 2^r$, that

$$(4.35) \qquad 2^{r-j} + j(m-x) < 2^r.$$

Since $m - x < 2^{r-j}$, it is sufficient to show that $j + 1 \leq 2^j$, and this clearly holds for all integers $j \geq 1$.

Since the powers of q on the left side of (4.33) (after dividing by $1 - q$) go up to q^{m-1}, we can cancel all the powers on the right from x to $m - 2g$ if $x \leq m - 2g$. We also cancel the powers from 0 to $x - 2f$ on both sides of (4.33). Then all the remaining powers on the left are at least x and those on the right are less than x. Hence (4.33) holds a fortiori, in view of the extra terms on the right.

Cases 2A and 2B with $f > 0$ and $x = m - 2^r + f$ are similar to the above and are also omitted. We now consider Case 1 with $f < 0$; Case 2 with $f < 0$ will also be omitted.

Suppose $x = 2^{r-1} + f$ with $-2^{r-2} \leq f < 0$ so that $2^{r-2} \leq x < 2^{r-1}$. We assume $m = 2^r + g$ where $0 \leq g < 2^{r-1}$ and Case 1 holds, i.e., $2^r \leq m < 3 \cdot 2^{r-1}$. Then $m - x = 2^{r-1} + g - f$. We consider two cases (1E and 1F) according as $0 \leq g - f < 2^{r-1}$ and $2^{r-1} \leq g - f < 3 \cdot 2^{r-2}$, respectively.

Under Case 1E we have $2^{r-1} \leq m - x < 2^r$ so that $r_{m-x} = r - 1$ and $d_{m-x} = 2^r - (m-x) = 2^{r-1} - g + f = x - g$. Also we have $r_x = r - 2$ and $d_x = -f$. To show that this x does not yield a maximum we compare the result with $f = 0$ in the computation of (4.16). We have to show that

$$(4.36)\qquad 1 - q^x + q^x(1-q^{x-g}) + 1 - q^{-f} \leq 1 - q^d.$$

Since $d = 2^{r+1} - m = 2^r - g = 2x - 2f - g$ this reduces to

$$1 - q^{-f} \leq q^{2x-g}(1-q^{-2f})$$

or (using the fact that $2x = m - g + 2f$)

$$(4.37)\qquad 1 \leq q^{m-2g+2f} + q^{m-2g+f}.$$

For this we use the fact that for $q > q_{m-1,m}$ we have $1 \leq q^{m-1} + q^m$. Hence it is sufficient to show that

$$(4.38)\qquad m - 2g + 2f \leq m \quad\text{and}\quad m - 2g + f \leq m - 1.$$

Both of these are immediate since $-f > 0$.

Under Case 1F, we have $2^r \leq m - x < 2^r + 2^{r-2}$ so that $r_{m-x} = r$, $d_{m-x} = 3\cdot 2^{r-1} - g + f = 3x - 2f - g = d + x$. Again $r_x = r - 2$ and $d_x = -f$. We have to show that for q in the interval $[q_{m-1,m}, q_{m,m+1}]$

$$(4.39)\qquad 1 - q^x - (q^x - q^m) + q^x(1-q^{d+x}) + 1 - q^{-f} \leq 1 - q^d$$

or, equivalently, since $d = m - 2g$ and $2x = m - g + 2f$

$$(4.40)\qquad 1 - q^x + q^{m-2g}(1 - q^{m-g+2f}) \leq q^{-f} - q^m.$$

Dividing by $1 - q$, we have to show that

$$(4.41)\qquad 1 + q + \ldots + q^{x-1} + q^{m-2g} + \ldots + q^{2m-3g+2f} \leq q^{-f} + \ldots + q^{m-1}.$$

For Case 1F $m < 3\cdot 2^{r-1} = \frac{m-g}{2} \Rightarrow g < \frac{m}{3}$. Also $g - f \geq 2^{r-1} = \frac{m-g}{2}$. This implies that $m + 2f \leq 3g < m$. Hence $m - 3g + 2f$ can be equal to m, m-1,... or $m + 2f + 1$. We consider the hardest case, where $2m - 3g + 2f = m$; the others follow similarly. Then $x = \frac{m-g}{2} + f = g$ and we can write (4.41) as

$$(4.42) \qquad 1 + q + \ldots + q^{-f-1} - q^{g} - q^{g+1} - \ldots - q^{g-2f-1} + q^{m} \leq 0.$$

This can be written as

$$(4.43) \qquad (1-q^{g}- q^{g+1} + q^{m}) + q(1-q^{g+1}- q^{g+2}) + \ldots + q^{-f-1}(1-q^{g-f-1}-q^{-f}) \leq 0.$$

We need only show that the first parenthesis in (4.43) is negative; the others are clearly negative since $m + 1 \geq 2^{r} > g - f$. For the first expression in (4.43) we divide by $1 - q$ and have to show that

$$(4.44) \qquad 1 + q + \ldots + q^{g-1}- q^{g+1} - \ldots - q^{m-1} \leq 0.$$

Since $m - 1 - g \geq 2g$, we can write (4.44) in the form

$$(4.45) \qquad (1-q^{g+1}- q^{g+2}) + q(1-q^{g+2}- q^{g+3}) + \ldots + q^{g-1}(1-q^{2g}- q^{2g+1}) - q^{3g+1} - \ldots - q^{m-1} \leq 0.$$

Since $2g < m - 1$ (this holds because $g > 0 > 1 + 2f$), it follows that every parenthesis in (4.45) is negative. This completes the proof for Case 1F and also completes the proof of Theorem 1.

5. Tables and Comparisons of Procedure R_T with other Procedures.

Table 1 gives the values of $U(c)$ and $U(c)/c$ for procedure R_T for $c = 1(1)25$, 100 and for $c = \infty$. It also gives the values $x_H(c)$ to be taken in any binomial (or H) situation that arises in the course of the experiment. For $c = \infty$ the values of $U(c)/c$ are taken from Table I of [5] and the values of $x_H(\infty)$ are taken from Table VII of [4]. The values of $x_H(c)$ are seen to converge to their limiting values $x_H(\infty)$ and in fact they appear to attain their limiting value at a value of c between $x_H(\infty)$ and $3x_H(\infty)$. Thus for $q = .99$ the limiting value $x_H(\infty) = 69$ is attained at $c = 79$, there is no change up to $c = 100$ and there appears to be no change thereafter.

Tables 2A, 2B, and 2C give the $x_G(m, c)$ values for $q = .90, .95$ and $.99$, respectively, for $c = 1(1)25, 100$ and for $c = \infty$. For $c = \infty$ the values are taken from Table IIIB of [4]. For $q = .90, .95$ and $.99$ and any $c \leq 25$, these tables (1 and 2) enable the investigator to carry out completely procedure R_T, as well as procedure $R_\infty = R_{2,1}$.

For each fixed $q > q_0 = (\sqrt{5} - 1)/2$ a table of $x_G(m, c)$ values is conjectured to have some regularity properties for $m \leq m(q)$, where $m(q)$ is the first value of m for which the given q lies in the interval $[q_{m-1,m}, q_{m,m+1}]$. One property is that the columns are monotonically nondecreasing in m for $m \leq m(q)$. Another is that $x_G(m, c)$ converges to its limiting value $x_G(m,\infty)$ and appears to attain this value at some $c \leq 3m/4$ for $m \leq m(q)$. A third property is that for $m \leq m(q)$ the rows are monotonically nondecreasing in c for $c \geq 1 + [\log_2(\frac{m+1}{3})]$. These properties have not been proved and should be treated as conjectures. For $m > m(q)$ the limiting properties established in Section 4 hold, but the approach is irregular.

Our asymptotic analysis in Section 4 enables us to make meaningful comparisons with other procedures by computing the asymptotic $(c \to \infty)$ efficiency of one procedure with respect to another. This efficiency will not depend on c or on the number N of units to be classified (we assume

$N = \infty$), but only on q. The efficiency of any procedure R with respect to R_∞ is defined by

$$(5.1) \qquad \mathrm{Eff}\{R/R_\infty\} = \frac{\lim_{c\to\infty} U_R(c)/c}{\lim_{c\to\infty} U(c)/c} = \lim_{c\to\infty} \frac{U_R(c)}{U(c)} = \frac{\lim_{N\to\infty} E\{T\ (N)/N\}}{\lim_{N\to\infty} E\{T_R(N)/N\}},$$

where $T_R(N)$ is the number of tests required to classify N units under procedure R and for procedure R_∞ the last numerator in (5.1) is the reciprocal of the $\lim U(c)/c$ as $c \to \infty$.

Consider the procedure R_1 defined in [4] for $N = 100$. Suppose we define a procedure R_1^* by dividing the units to be classified into groups of 100 each and we use procedure R_1 for each group of 100. For $q = .99$ we can use an entry in Table I of [5] (see remark after (4.21)) and an entry in Table Vc of [4] to obtain for the asymptotic efficiency

$$(5.2) \qquad \mathrm{Eff}(R_1^*/R_\infty;\ q = .99) = \frac{.08105}{.08320} = 97.42 \text{ percent.}$$

We use the same basic equations (2.1) through (2.5) to define $R_\infty(c)$, except that instead of taking the maximum on the right side of (2.1) and (2.2) we use the $x_H(q)$-values and the $x_G(m)$-values from the procedure $R_\infty = R_{2,1}$. Since R_∞ is the limit of R_T the asymptotic $(c \to \infty)$ efficiency of $R_\infty(c)$ with respect to $R_T(c)$ is 100 percent for all values of q. The corresponding efficiencies for fixed values of c are given in Table 3 for $q = .90$, $.95$ and $.99$.

To define procedure $R_2(c)$ we again use the same basic equations as above but we take the $x_H(q)$-values and $x_G(m, q)$-values from the information procedure R_2, defined in the appendix of [4]. For $q = .90$, $.95$ and $.99$ the efficiencies for $R_2(c)$ with respect to $R_T(c)$ are given in Table 3 for finite c and for $c = \infty$.

The limit of $R_2(c)$ as $c \to \infty$ is R_2. We can obtain the asymptotic ($c \to \infty$) efficiency of $R_2(c)$ (or the efficiency of R_2) relative to R_∞ by assuming that $U(c|R_2(c)) \sim Bc$ and computing the value of $B = B(q)$ as was done in Section 4 for procedure R_T. The ratio of this B to the b-value in (4.18) is the desired asymptotic efficiency. Thus for $.8899 < q < .8987$ (cf. Table VII of [4]) we take $x = 6$ and for $m = 6$ in the G-situation we take $x = 3$. To derive the corresponding B-value, we write (using U_2 for procedure R_2)

$$(5.3) \qquad U_2(c) = q^6[6 + U_2(c-1)] + (1-q^6)U_2(6,\ c-1)$$

$$(5.4) \qquad U_2(6,\ c-1) = \frac{3(q^3 - q^6)}{1 - q^6} + U_2(3,\ c-2) = \frac{1 + 2q + \ldots + 6q^5 + B(1+q^3)}{1 + q + \ldots + q^5} + B(c-4).$$

Substituting (5.4) into (5.3) and writing Bc for $U_2(c)$ and $B(c-1)$ for $U_2(c-1)$, we find that

$$(5.5) \qquad B = \frac{1 - q^6}{p[3(1-q^6) + q(1+q^3-q^2)]} .$$

Using (4.18) with $x = 6$, $r_x + 1 = 3$ and $d_x = 1$, we obtain

$$(5.6) \qquad \mathrm{Eff}\{R_2/R_\infty\} = \frac{3(1-q^6) + q^2}{3(1-q^6) + q(1-q^2+q^3)} \quad \text{for} \quad .8899 < q < .8987.$$

Some expressions obtained for other ranges of q are

$$(5.7) \qquad \mathrm{Eff}\{R_2/R_\infty\} = \frac{4(1-q^{14}) + q^2}{4(1-q^{14}) + q(1-q^2+q^3)} \quad \text{for} \quad .9499 < q < .9533,$$

$$(5.8) \qquad \mathrm{Eff}\{R_2/R_\infty\} = \frac{7(1-q^{69}) + q^{59}}{7 - 6q^{69} + (1-q)(1+q^3+q^6)[q^{50}(1+q^{11})(1+q)-1]+q^{36}(1-q^2)(1+q^9)}$$

where the latter holds for $.9899 < q < .9901$. A list of the numerical efficiencies thus obtained for the discrete set of values $q = .01(.01).99$ is given in Table 4. Although the results are not monotonic in q, they are all greater than 99 percent and it becomes plausible to conjecture that this lower bound (.99) holds uniformly for all values of q.

In the class of procedures that always break down the defective set when it is present (without mixing in any binomial units) the procedure R_T is optimal; this is clear from the equations (2.1) through (2.5). If $\text{Eff}\{R_2/R_\infty\} \geq .99$ for all q then we can regard procedure R_2 as an ϵ-admissible procedure in this class for $\epsilon = .01$.

Lest the reader think that the asymptotic efficiency is always high, we now compute the asymptotic efficiency of the original Dorfman procedure R_D [1]. For $q = .99$ we take 11 units and if the pooled result is negative, we test them one-at-a-time. Hence for $q = .99$ under procedure R_D

$$(5.9) \qquad E\{T|R_D, q = .99\} = q^{11} + 12(1-q^{11}) = 2.1513,$$

and the expected number of units classified per test is 11/2.1513 or 5.1132. Since the corresponding value for R_∞ is 12.338, we have for the asymptotic efficiency of R_D with respect to R_∞

$$(5.10) \qquad \text{Eff}\{R_D/R_\infty; q = .99\} = \frac{5.1132}{12.338} = 41.44 \text{ percent.}$$

On the basis of this, we regard R_D as an inefficient procedure.

Sterrett [8] made a slight improvement to the Dorfman procedure by testing units in a defective set one-at-a-time only until a defective unit is found. Then the remaining units (from the original batch of $n = 15$ for $q = .99$) are tested as a group and either passed in 1 test or again tested individually. If we start with

n units then this procedure R_S leads to an expected number of tests $H_S(n)$ for n units given by

$$(5.11) \qquad H_S(n) = q^n + \sum_{i=1}^{n-1} pq^{i-1}[i+1+H_S(n-i)] + npq^{n-1}.$$

Here inference is used (in obtaining the last term) to save one test, if the last unit is defective and the next-to-last is good. For any $n \geq 1$ and any q, if we add the expressions for $pH_S(j)$ $(j = 1, 2, \ldots, n-1)$ to (5.11), we obtain

$$(5.12) \qquad H_S(n) = n + 1 + (n-2)p - \frac{q}{p}(1-q^n).$$

This exact expression (5.12) was not obtained in [8] (cf. eq. (160) of [4]). Contrary to the instruction in [8], we work with groups of size $n = 15$ to minimize $H_S(n)/n$ when $q = .99$. Using (5.12) this gives $H_S(15)/15 = .15172$; the larger value of $H_S(16)/16$ which is .15215 is incorrectly given in [8] as .14. Hence the efficiency of R_S with respect to R_∞ for $q = .99$ is

$$(5.13) \qquad \mathrm{Eff}(R_S/R_\infty;\ q = .99) = \frac{.08105}{.15172} = 53.42 \text{ percent.}$$

Another improvement and extension of Dorfman's work was made by Finucan in [2] without being aware of other work on the same topic. "He tests all his units in disjoint groups of equal (or approximately equal) size, say x_1, and this constitutes his first stage. He then tests all groups showing at least 1 defective (with and without recombining defective sets after the first stage) using disjoint subgroups of equal (or approximately equal) size, say $x_2 < x_1$, and this is his second stage. If he tests individual units the j^{th} time around, then it is called a j-stage procedure." If we substitute $r = 1/p = 100$ in his equation (7) for the minimum expected number of tests per unit classified (using the optimal number of stages), we obtain $c_{min} = (.02718)(4.60517) = .12517$. This is comparable with the result .08320 for $N = 100$, $p = .01$ in Table Vc of [4] and with .08105 for $p = .01$ under procedure $R_\infty \doteq R_{2,1}$. Hence the efficiency of his procedure R_F with respect to R_∞ for $q = .99$ is

$$(5.14) \qquad \mathrm{Eff}(R_F/R_\infty;\ q = .99) = \frac{.08105}{.12517} = 64.75 \text{ percent.}$$

Thus, taking 90 percent as a minimal acceptable standard, we find that the optimal procedure using this scheme is still inefficient.

The reader might at this point ask whether <u>all</u> of our procedures are in fact inefficient since we have not shown that R_T or R_∞ are optimal procedures or even asymptotically optimal in the class of all group-testing procedures. To answer this we point out some known results on how <u>close</u> the procedures R_T and R_∞ are to being optimal in the class of all group-testing procedures. A lower bound to the expected number of tests per unit classified is $-[p \log_2 p + q \log_2 q]$ which is equal to .08079 for $q = .99$. The reciprocal, 12.378, is then an upper bound (for $q = .99$) to the expected number of units classified per test; these bounds hold for <u>all</u> group-testing procedures. Thus for $q = .99$ our measure of asymptotic efficiency always overestimates and can differ from the absolute measure of efficiency, i.e., a measure which compares each procedure with the optimal procedure, by at most the factor $.08079/.08105 = .9968$. In other words, for $q = .99$ the absolute efficiency is less than our efficiency (relative to R_T or $R_\infty = R_{2,1}$) and differs from it by a negligible amount (at most about 1/3 of 1 percent).

Attempts to improve R_T or $R_\infty = R_{2,1}$ can be made corresponding to the attempts to improve R_1 in [5] and [7]. It follows from the above discussion that any improvement will at most be slight and at the cost of complications in the instructions required to carry out the more efficient procedure. There is also a possibility that no improvement is possible for procedure $R_{2,1}$ with $N = \infty$; this has not yet been investigated.

6. Acknowledgement.

The authors wish to thank Professor William D. Sudderth of the University of Minnesota for making some key suggestions that led to the asymptotic analysis in Section 4. We also wish to acknowledge the assistance of Mr. Leo May for programming the procedure R_T, the results of which were used in almost all the tables of this paper.

Table 1: Values of U(c), U(c)/c and $x_H(c)$ for Procedure R_T as a Function of c for q = .90, .95 and .99.

c	q = .90			q = .95			q = .99		
	U(c)	U(c)/c	$x_H(c)$	U(c)	U(c)/c	$x_H(c)$	U(c)	U(c)/c	$x_H(c)$
1	3.487	3.487	9	7.170	7.170	20	36.603	36.603	99
2	5.932	2.966	10	12.201	6.100	21	62.278	31.139	110
3	7.592	2.531	8	15.405	5.135	23	78.648	26.216	118
4	9.620	2.405	4	17.706	4.426	11	88.288	22.072	127
5	11.896	2.379	6	20.891	4.178	8	93.717	18.743	107
6	14.195	2.366	8	24.743	4.124	10	100.088	16.681	16
7	16.172	2.310	6	28.919	4.131	16	110.130	15.733	32
8	18.220	2.277	6	32.550	4.069	16	124.476	15.559	34
9	20.379	2.264	6	35.529	3.948	15	142.409	15.823	64
10	22.568	2.258	7	38.712	3.871	11	159.690	15.969	74
11	24.653	2.241	7	42.259	3.842	12	174.077	15.825	98
12	26.734	2.228	6	46.025	3.835	14	184.484	15.374	112
13	28.858	2.220	6	49.678	3.821	16	192.134	14.780	71
14	31.002	2.214	7	52.986	3.785	14	201.179	14.370	43
15	33.115	2.208	7	56.279	3.752	13	212.772	14.185	45
16	35.216	2.201	7	59.726	3.733	13	226.876	14.180	59
17	37.330	2.196	7	63.313	3.724	13	242.443	14.261	64
18	39.457	2.192	7	66.896	3.716	14	257.518	14.307	73
19	41.575	2.188	7	70.337	3.702	14	270.685	14.247	87
20	43.686	2.184	7	73.720	3.686	13	281.334	14.067	93
21	45.800	2.181	7	77.155	3.674	13	290.822	13.849	68
22	47.920	2.178	7	80.663	3.666	13	301.364	13.698	54
23	50.038	2.176	7	84.192	3.661	14	313.547	13.632	57
24	52.152	2.173	7	87.670	3.653	14	327.255	13.636	64
25	54.267	2.171	7	91.105	3.644	14	341.612	13.664	64
50	107.177	2.144	7	177.887	3.558	14	650.126	13.003	72
100	212.995	2.130	7	351.450	3.515	14	1266.423	12.664	69
∞	∞	2.116	7	∞	3.466	14	∞	12.338	69

Table 2: **Values[§] of $x_G(m, c)$ to be used with Procedure R_T for three Values of q.**

Table 2A: $q = .90$

m \ c	∞	100	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	c / m
2	1	1																									1	2
3	1	1																									1	3
4	2	2																							2	1	1	4
6	2	2																							2	1	3	6
7	3	3																						3	2	1	3	7
8	4	4										4	3	4	4	3	3	3	4	4	3	3	4	4	2	1	4	8
10	4	4																						4	2	1	4	10

Table 2B: $q = .95$

m \ c	∞	100	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	c / m
2	1	1																									1	2
3	1	1																									1	3
4	2	2																							2	1	1	4
5	2	2																							2	1	2	5
6	2	2																							2	1	3	6
7	3	3																						3	2	1	3	7
8	4	4																						4	2	1	4	8
9	4	4																						4	2	1	4	9
10	4	4																						4	2	1	5	10
11	4	4																						4	2	1	5	11
12	4	4																						4	2	6	6	12
13	5	5																	5	4	5	5	5	4	2	6	6	13
14	6	6																6	5	4	6	6	6	4	2	6	6	14
15	7	7										7	6	6	6	7	7	6	6	5	7	7	7	4	2	7	7	15
16	7	7								7	8	7	7	6	7	8	8	7	6	5	8	8	8	4	2	7	7	16
21	8	8																8	7	7	8	8	8	4	2	9	9	21
23	8	8																8	7	7	8	8	8	4	2	10	10	23

[§] The bold line indicates where the entries "settle down" and become equal to the row values for $c = \infty$.

Table 2C: q = .99

m \ c	∞	100	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	c
2	1	1																									1	2
3	1	1																									1	3
4	2	2																							2	1	1	4
5	2	2																							2	1	2	5
6	2	2																							2	1	3	6
7	3	3																						3	2	1	3	7
8	4	4																						4	2	1	4	8
9	4	4																						4	2	1	4	9
10	4	4																						4	2	1	5	10
11	4	4																						4	2	1	5	11
12	4	4																						4	2	1	6	12
13	5	5																					5	4	2	1	6	13
14	6	6																					6	4	2	1	7	14
15	7	7																					7	4	2	1	7	15
16	8	8																					8	4	2	1	8	16
17	8	8																					8	4	2	1	8	17
18	8	8																					8	4	2	1	9	18
19	8	8																					8	4	2	1	9	19
20	8	8																					8	4	2	1	10	20
21	8	8																					8	4	2	10	10	21
22	8	8																					8	4	2	10	11	22
23	8	8																					8	4	2	11	11	23
24	8	8																					8	4	2	12	12	24
25	9	9																				9	8	4	2	12	12	25
26	10	10																				10	8	4	2	12	13	26
27	11	11																				11	8	4	2	13	13	27
28	12	12																				12	8	4	2	14	13	28
29	13	13														13	12	13	13	13	13	13	8	4	2	14	14	29
30	14	14														14	13	14	14	14	14	14	8	4	2	14	14	30
31	15	15														15	13	15	15	15	15	15	8	4	2	15	15	31
32	16	16														16	13	16	16	16	16	16	8	4	2	16	16	32
33	16	16														16	14	16	16	16	16	16	8	4	2	16	16	33
34	16	16														16	14	16	16	16	16	16	8	4	2	16	16	34

cont. cont.

Table 2C: (cont.)

m \ c	∞#	100	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	c / m
40	16	16																				16	8	4	13	19	19	40
41	16	16																				16	8	4	13	19	19	41
42	16	16																				16	8	4	14	20	20	42
43	16	16																				16	8	4	14	20	20	43
45	16	16																				16	8	4	15	21	21	45
47	16	16																				16	8	4	15	22	22	47
48	16	16																				16	8	4	15	23	22	48
50	18	18												18	17	16	18	18	18	18	18	16	8	4	16	23	23	50
54	22	22												22	18	16	22	22	22	22	22	16	8	4	25	25	25	54
55	23	23					23	22	23	23	23	23	23	23	19	16	23	23	23	23	23	16	8	4	33	26	26	55
57	25	25				25	23	25	25	25	25	25	24	19	16	25	25	25	25	25	25	16	8	4	27	26	26	57
59	27	27				27	25	24	26	27	27	27	27	25	20	17	27	27	27	27	27	16	8	4	28	27	27	59
61	29	29	29	29	29	29	26	24	27	29	29	29	29	25	21	18	27	29	29	29	29	16	8	4	28	28	28	61
63	31	31	31	31	31	30	27	25	28	31	31	31	31	26	21	18	28	31	31	31	31	16	8	4	29	29	29	63
64	32	32	32	32	32	30	27	25	28	32	32	32	32	26	22	19	28	32	32	32	24	16	8	4	30	29	29	64
68	32	32	32	32	32	32	28	27	29	32	32	32	32	28	23	20	30	32	32	32	26	16	8	4	31	31	31	68
69	32	32	32	32	32	32	29	27	30	32	32	32	32	28	23	20	30	32	32	32	26	16	8	4	31	31	31	69
71	32	32	32	32	32	32	29	28	30	32	32	32	32	29	23	21	31	32	32	32	27	16	8	4	33	33	32	71
73	32	32				32	30	28	31	32	32	32	32	29	24	22	31	32	32	32	27	16	8	4	33	33	33	73
74	32	32				32	30	29	31	32	32	32	32	30	24	22	32	32	32	32	28	16	8	4	34	34	33	74
76	32	32				32	31	29	32	32	32	32	32	30	24	23	32	32	32	32	28	16	8	15	34	34	34	76
87	32	32												32	26	26	32	32	32	32	28	16	8	17	38	38	38	87
93	32	32												32	26	26	32	32	32	32	29	16	8	18	41	41	41	93
98	34	34	34	34	32	32	32	32	34	34	34	34	32	32	27	27	34	34	34	32	27	16	8	29	43	43	43	98
107	U	37	42	38	34	32	32	32	43	43	43	39	32	32	28	29	43	43	43	32	29	16	8	31	47	46	46	107
110	U	38	43	39	35	32	32	33	46	46	44	40	32	32	29	31	46	46	46	32	28	16	8	32	47	47	47	110
112	U	39	43	40	36	32	32	33	48	48	45	40	32	32	29	32	48	48	48	32	28	16	8	32	48	48	47	112
118	U	41	45	42	38	34	33	35	51	52	47	42	33	32	30	34	54	54	54	32	28	16	8	43	50	50	50	118
127	U	44	48	45	41	36	35	38	54	55	50	45	34	32	30	38	63	62	59	32	29	16	8	46	54	53	52	127

#For $c = \infty$, the values up to $m = 100$ are obtainable from Page 69 of [3]. The five entries marked U have not been computed. According to a conjecture in Section 5 they should be equal to the value at $c = 100$ for all $m \leq 133$. Thus for $m = 98$ and 127 the entries appear to settle down at $c = 54$ (to the value $x = 34$) and at $c = 66$ (to the value $x = 44$), respectively.

Table 3: Efficiency of $R_\infty(c)$ and $R_2(c)$ Relative to $R_T(c)$ for Selected Values of c.

c	Procedure $R_\infty(c)$			Procedure $R_2(c)$		
	$q = .90^*$	$q = .95$	$q = .99$	$q = .90^*$	$q = .95$	$q = .99$
1	96.02	95.23	94.22	96.02	95.23	94.22
2	96.11	95.42	94.63	96.11	95.42	94.58
3	96.45	95.87	95.43	96.45	96.08	95.39
4	94.12	95.06	96.43	94.12	95.11	96.43
5	96.92	91.52	97.35	96.92	91.39	97.36
10	98.36	96.39	96.62	98.36	96.28	96.20
25	99.27	98.73	97.81	99.27	98.60	97.48
50	99.63	99.34	98.87	99.63	99.20	98.61
100	99.82	99.67	99.39	99.82	99.52	99.14
∞	100.00	100.00	100.00	100.00	99.84	99.73

*These two procedures are the same at $q = .90$ and hence these columns are identical.

Table 4: Asymptotic Efficiency of $R_2(c)$ Relative to R_∞ for $c \to \infty$.

q	$\mathrm{Eff}(R_2/R_\infty)$ in %	$x_H(R_2)^\S$	$x_G(x_H, q \mid R_2)^\S$
0 (.01).81	100	≤ 3	1
.82(.01).85	100	4	2
.86(.01).88	100	5	2
.89	99.12	6	3
.90	100	7	3
.91	100	7	3
.92	99.47	8	3
.93	99.76	10	4
.94	99.41	11	5
.95	99.84	14	6
.96	99.74	17	7
.97	99.49	23	10
.98	99.65	34	14
.99	99.73	69	29

§All these values are taken from Table VII of [3]. Column 4 is the particular value of $x_G(m, q \mid R_2)$ when $m = x_H(R_2)$; the latter is given in column 3.

References

[1] Dorfman, R. (1943). The detection of defective members of large populations. Ann. Math. Stat. 14 436.

[2] Finucan, H. M. (1964). The blood testing problem. Appl. Statist. 13 43-50.

[3] Kumar, S. (1970). Multinomial group-testing. SIAM Jour. App. Math. 19.

[4] Sobel, M. and Groll, P. A. (1959). Group-testing to eliminate efficiently all defectives in a binomial sample. Bell System Tech. Jour. 38 1179-1252.

[5] Sobel, M. (1960). Group-testing to classify all defectives in a binomial sample. Information and Decision Processes, ed. R. E. Machol. McGraw-Hill, 127-161.

[6] Sobel, M. (1968). Binomial and Hygeometric group-testing. Studia Sci. Math. Hungar. 3 19-42.

[7] Sobel, M. (1967). Optimal group-testing. Proceedings of the Colloquium on Information Theory Organized by the Bolyai Mathematical Society, Debrecen, (Hungary), 411-488.

[8] Sterrett, A. (1957). On the detection of defective members of large populations. Ann. Math. Stat. 28 1033.